Lecture 14: Planted Sparse Vector

Lecture Outline

- Part I: Planted Sparse Vector and 2 to 4 Norm
- Part II: SOS and 2 to 4 Norm on Random Subspaces
- Part III: Warmup: Showing $||x|| \approx 1$
- Part IV: 4-Norm Analysis
- Part V: SOS-symmetry to the Rescue
- Part VI: Observations and Loose Ends
- Part VII: Open Problems

Part I: Planted Sparse Vector and 2 to 4 Norm

Planted Sparse Vector

- Planted Sparse Vector problem: Given the span of $d-1$ random vectors in $\mathbb{R}^n$ and one unit vector $v \in \mathbb{R}^n$ of sparsity k, can we recover v?
- More precisely, let V be an $\mathrm{n} \times d$ matrix where:
  1. $d-1$ columns of V are vectors of length ≈ 1 chosen randomly from $\mathbb{R}^n$
  2. One column of V is a unit vector v with $\leq k$ nonzero entries.
- Given VR where R is an arbitrary invertible $d \times d$ matrix, can we recover v?

Theorem Statement

- Theorem 1.4 [BKS14]: There is a constant $c > 0$ and an algorithm based on constant degree SOS such that for every vector v_0 supported on at most $cn \cdot \min\{1, n/d^2\}$ coordinates, if $v_1, \ldots, v_d$ are chosen independently at random from the Gaussian distribution on R^n, then given any basis for $V = span\{v_0, \ldots, v_d\}$, the algorithm outputs an ϵ-approximation to v_0 in $poly(n, \log(1/\epsilon))$ time.

Random Distribution

- Random Distribution: We choose each entry of V independently from $N\left(0, \frac{1}{n}\right)$, the normal distribution with mean 0 and standard deviation $\frac{1}{\sqrt{n}}$
- We then choose R to be a random $d \times d$ orthogonal/rotation matrix and take VR to be our input matrix.

Random Distribution

- Remark: If R is any $d \times d$ orthogonal/rotation matrix then VR can also be chosen by taking each entry of V independently from $N\left(0, \frac{1}{n}\right)$.
- Idea: Each row of V comes from a multivariate normal distribution with covariance matrix $\frac{1}{n} Id_d$, which is invariant under rotations

Planted Distribution

- Planted Distribution: We choose each entry of the first $d-1$ columns of V independently from $N\left(0, \frac{1}{n}\right)$. The last column of V is our sparse unit vector v.
- We then choose R to be a random $d \times d$ orthogonal/rotation matrix and take VR to be our input matrix.

Output

- We ask for an x such that
  1. $\|VRx\| = 1$
  2. VRx is k-sparse (i.e. at most k indices of VRx are nonzero).
- Hard to search for x such that VRx is k-sparse, so we'll need to relax the problem.

Distinguishing Sparse Vectors

- Key idea: All unit vectors have the same 2-norm. However, sparse vectors will have higher 4-norm
- 4-norm for a k-sparse unit vector in $\mathbb{R}^n$ is at least $\sqrt[4]{\mathrm{k} \cdot \frac{1}{k^2}} = \frac{1}{\sqrt[4]{k}}$ (obtained by setting k coordinates to $\frac{\pm 1}{\sqrt{k}}$ and the rest to 0)
- Relaxation Attempt #1: Search for an x such that
  1. $\|VRx\| = 1$
  2. $\|VRx\|_4 \geq \frac{1}{\sqrt[4]{k}}$

2 to 4 Norm Problem

- This is the 2 to 4 Norm Problem: Given a matrix A, find the vector x which maximizes $\frac{\|Ax\|_4}{\|Ax\|}$

Part II: SOS and 2 to 4 Norm on Random Subspaces

2 to 4 Norm Hardness

- Unfortunately, the 2 to 4 norm problem is hard [BBH+12]:
  - NP-hard to obtain an approximation ratio of $\left(1+\frac{1}{npolylog(n)}\right)$
  - Assuming ETH (the exponential time hypothesis), it is hard to approximate to within a constant factor.
- Thus, we’ll need to relax our problem further.

SOS Relaxation

- Relaxation: Find $\tilde{E}$ which respects the following constraints:
  1. $\|VRx\|^2 = \sum_{i=1}^{n}(VRx)_i^2 = 1$
  2. $\|VRx\|_4^4 = \sum_{i=1}^{n}(VRx)_i^4 \geq \frac{1}{k}$

Showing a Distinguishing Algorithm

- Constraints:
  1. $\|VRx\|^2 = \sum_{i=1}^{n}(VRx)_i^2 = 1$
  2. $\|VRx\|_4^4 = \sum_{i=1}^{n}(VRx)_i^4 \geq \frac{1}{k}$
- To show that SOS distinguishes between the random and planted distribution, it is sufficient to show that there is no $\tilde{E}$ which respects these constraints and has a PSD moment matrix M.
- Remark: Although the 2 to 4 Norm problem is hard in general, we just need to show that SOS can approximate it on random subspaces.

2 to 4 Norm on Random Subspaces

- Given a random subspace, what is the expected value of the largest 4-norm of a unit vector in the subspace?
- Trivial strategy: Any unit vector's 4-norm is at least $\frac{1}{\sqrt[4]{n}}$.
- Can we do better?

2 to 4 Norm on Random Subspaces

- Another strategy: Take a basis for this space and take a linear combination which maximizes one coordinate (subject to having length 1)
- If we add together d random vectors with entries $\approx \pm\frac{1}{\sqrt{n}}$, w.h.p. the result will have norm $\widetilde{\Theta}\left(\sqrt{d}\right)$. Diving the resulting vector by $\widetilde{\Theta}\left(\sqrt{d}\right)$, the maximized entry will have magnitude $\widetilde{\Theta}\left(\frac{\sqrt{d}}{\sqrt{n}}\right)$, other entries will have magnitude $\widetilde{O}\left(\frac{1}{\sqrt{n}}\right)$

2 to 4 Norm on Random Subspaces

- Calling our final result w, w.h.p. the maximized entry of w contributes $\tilde{\Theta}\left(\frac{d^2}{n^2}\right)$ to $\|w\|_4^4$ while the other entries contribute $\tilde{\Theta}\left(\frac{1}{n}\right)$.
- It turns out that this strategy is essentially optimal. Thus, with high probability the maximum 4-norm of a unit vector in a d-dimensional random subspace will be $\tilde{\Theta}\left(\max\left\{\frac{\sqrt{d}}{\sqrt{n}}, \frac{1}{\sqrt[4]{n}}\right\}\right)$.

Algorithm Boundary

- Planted dist: max 4-norm $\geq \frac{1}{\sqrt[4]{k}}$
- Random dist: max 4-norm is $\widetilde{\Theta}\left(\max\left\{\frac{\sqrt{d}}{\sqrt{n}}, \frac{1}{\sqrt[4]{n}}\right\}\right)$.
- **IF** SOS can certify the upper bound for a random subspace, this gives a distinguishing algorithm when $\max\left\{\frac{\sqrt{d}}{\sqrt{n}}, \frac{1}{\sqrt[4]{n}}\right\} \ll \frac{1}{\sqrt[4]{k}}$ (which happens when $d \leq \sqrt{n}$ and $k \ll n$ or when $d \geq \sqrt{n}$ and $\mathrm{k} \ll \frac{n^2}{d^2}$)

Part III: Warmup: Showing $\|x\| \approx 1$

Showing $\|x\| \approx 1$

- Take $w = VRx$.
- We expect that $\|w\| \approx \|x\|$. Since we require that $\|w\| = 1$, this implies that we will have $\|x\| \approx 1$
- To check that $\|w\| \approx \|x\|$, observe that $\|w\|_2^2 = x^T(\mathrm{RV})^{\mathrm{T}}(\mathrm{VR})\mathrm{x}$. Thus, it is sufficient to show that $(\mathrm{RV})^{\mathrm{T}}(\mathrm{VR}) \approx Id$.

Checking $(\mathrm{RV})^{\mathrm{T}}(\mathrm{VR}) \approx Id$

- We have that $(\mathrm{RV})^{\mathrm{T}}(\mathrm{VR}) \approx Id$ because the columns of VR are d random unit vectors (where $d \ll n$) and are thus approximately orthonormal.
- However, we will use graph matrices to analyze the 4-norm, so as a warm-up, let's check that $(\mathrm{RV})^{\mathrm{T}}(\mathrm{VR}) \approx Id$ using graph matrices.

Graph Matrices Over $N(0,1)$

- So far we have worked over $\{-1,+1\}^m$.
- How can we use graph matrices over $N(0,1)^m$?
- Key idea: Look at the Fourier characters over $N(0,1)$.

Fourier Analysis Over $N(0,1)$

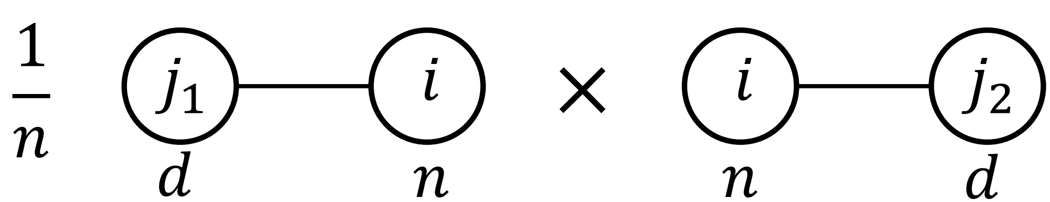
- Inner product on $N(0,1)$: $f \cdot g = E_{x\sim N(0,1)} f(x)g(x)$
- Fourier characters: Hermite polynomials
- The first few Hermite polynomials (up to normalization) are as follows:
  1. $h_0 = 1$
  2. $h_1 = x$
  3. $h_2 = x^2 - 1$
  4. $h_3 = x^3 - 3x$
- To normalize, divide h_j by $\sqrt{j!}$

Graph Matrices Over $N(0,1)$

- Graph matrices over $\{-1,1\}^m$: 1 and x are a basis for functions over $\{-1,1\}$. We represent x by an edge and 1 by the absence of an edge
- Graph matrices over $N(0,1)^m$: $\{h_j\}$ are a basis for functions over $N(0,1)$. We represent h_j by a multi-edge with multiplicity j.

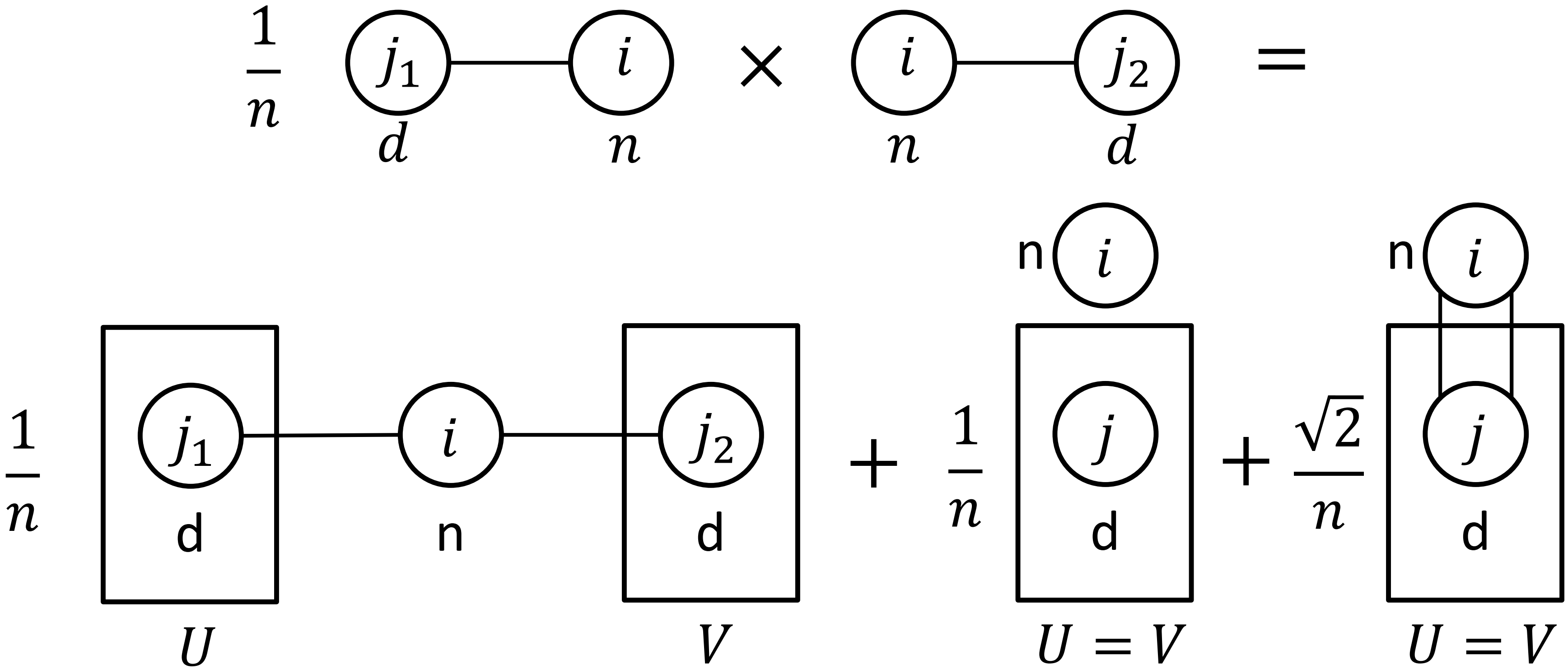
Graph Matrices for $(\mathrm{RV})^{\mathrm{T}}(\mathrm{VR})$

- For convenience, take $A = \sqrt{n}RV$ and think of the entries of A as the input. Now each entry of A is chosen independently from $N(0,1)$
- A_{ij} is represented by an edge from node i to node j.
- In class challenge: What is $(\mathrm{RV})^{\mathrm{T}}(\mathrm{VR})$ in terms of graph matrices?



Graph Matrices for $(\mathrm{RV})^{\mathrm{T}}(\mathrm{VR})$

- In class challenge answer:



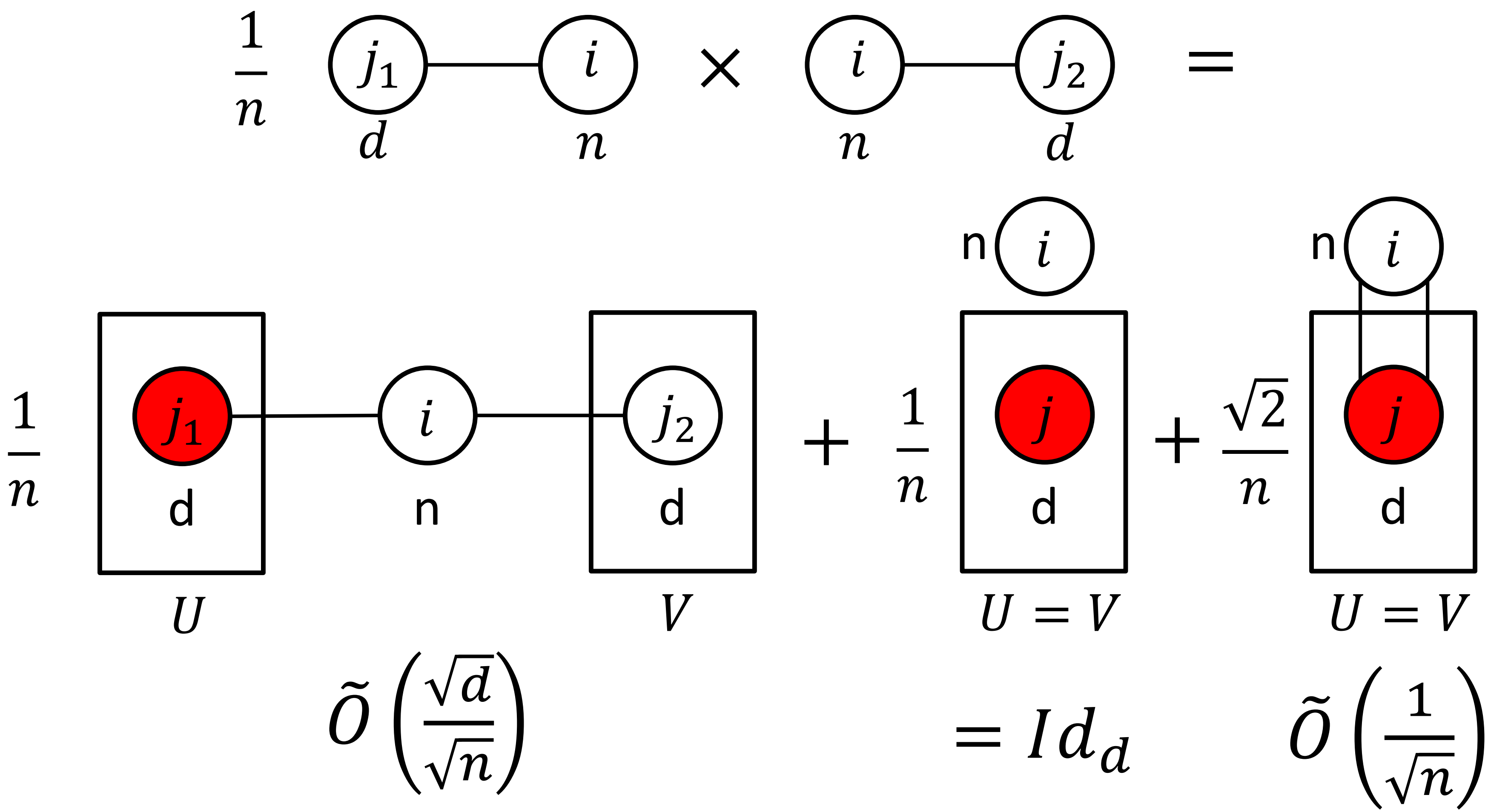
Generalizing Rough Norm Bounds

- Here we have two different types of vertices, one for the rows of A (which has n possibilities) and one for the columns of A (which has d possibilities)
- Can generalize the rough norm bounds to handle multiple types of vertices (writing this up is on my to-do list)

Generalizing Rough Norm Bounds

- Generalized rough norm bounds:
- Each isolated vertex outside of U and V contributes a factor equal to the number of possibilities for that vertex
- Each vertex in the minimum separator (which minimizes the total number of possibilities for its vertices) contributes nothing
- Each other vertex contributes a factor equal to the square root of the number of possibilities for that vertex

Norm Bounds for $(\mathrm{RV})^{\mathrm{T}}(\mathrm{VR})$



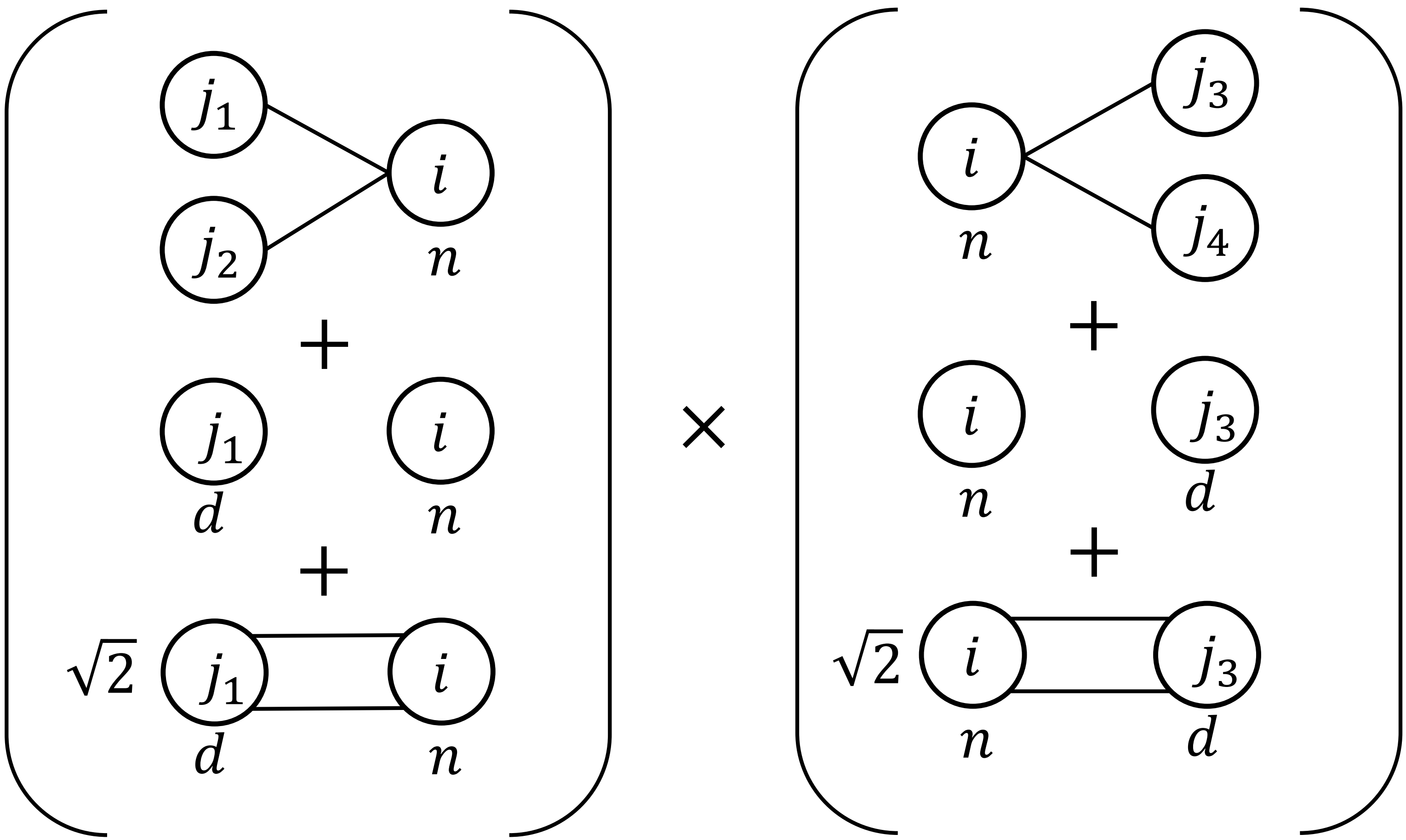
Part IV: 4-Norm Analysis

4-Norm Analysis

- We want to bound $\left\|\frac{1}{\sqrt{n}}Ax\right\|_4^4$
- Take B to be the matrix with entries $B_{i,(j_1,j_2)} = A_{ij_1}A_{ij_2}$
- $\left\|\frac{1}{\sqrt{n}}Ax\right\|_4^4 = \frac{1}{n^2}(x \otimes x)^T B^T B(x \otimes x)$
- Can try to bound $\left\|B^T B\right\|$

Picture for $B^T B$

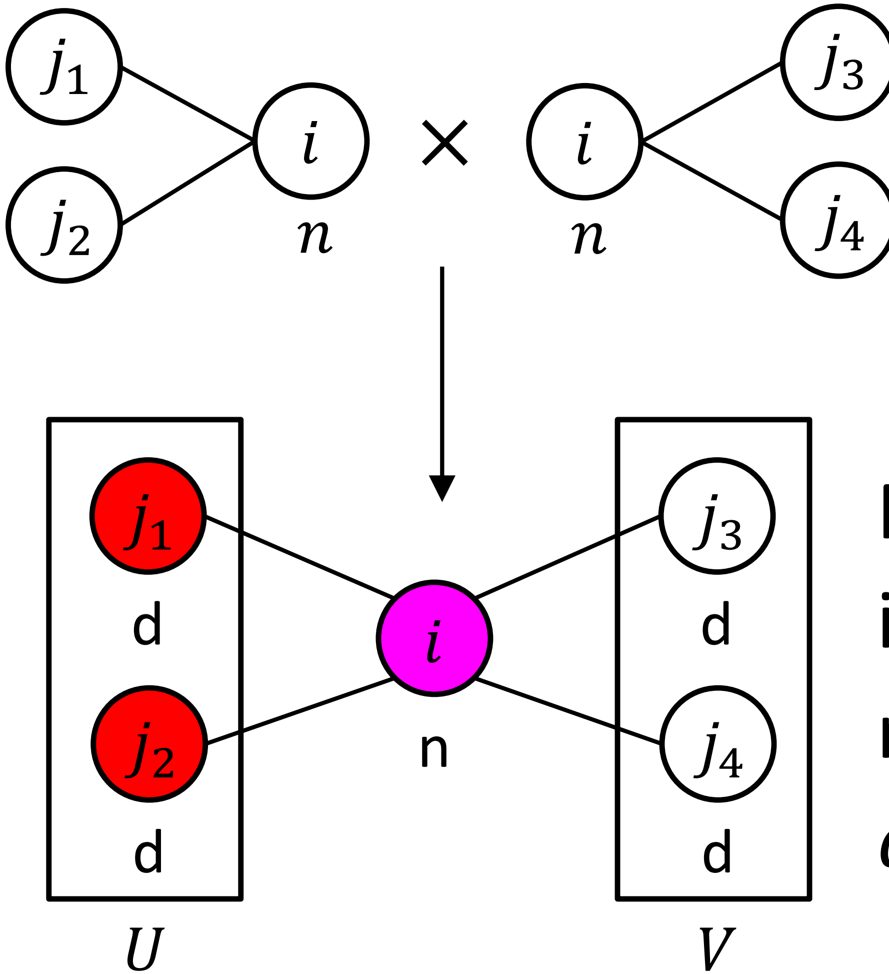
- Picture for $B^T B$:



Targets

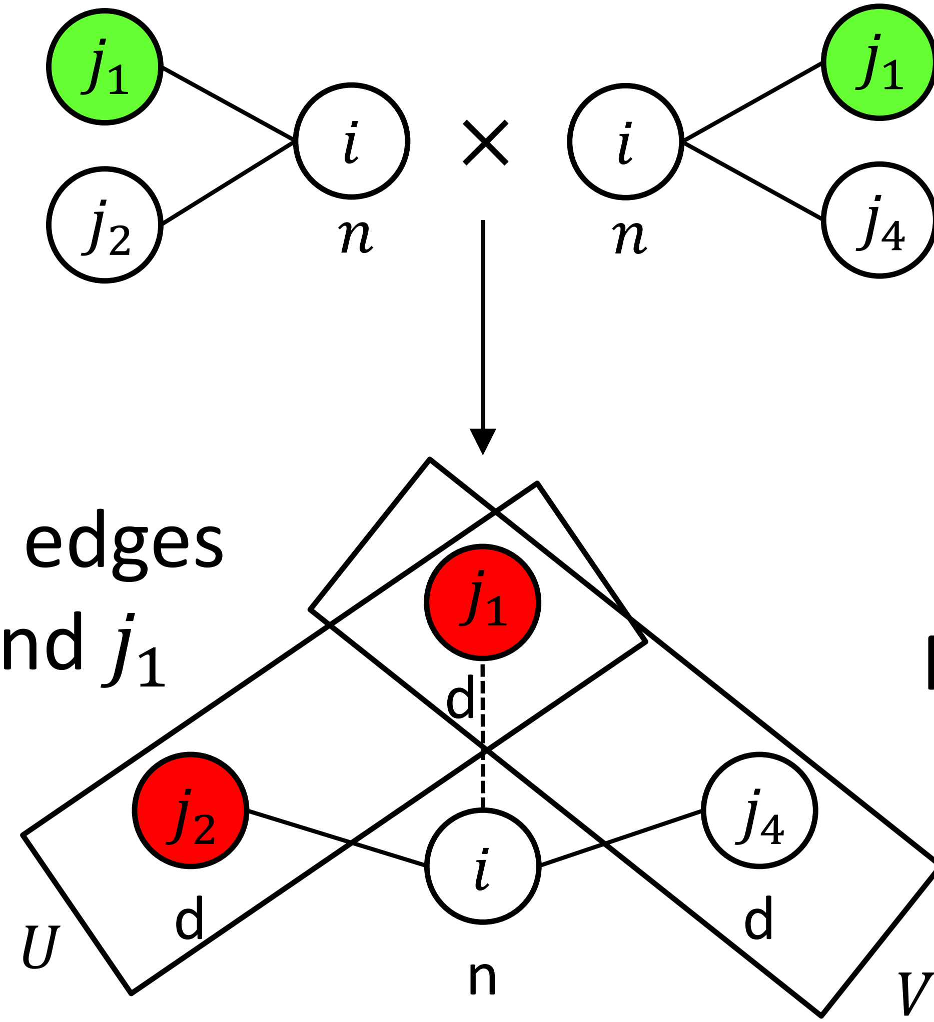
- If $d \leq \sqrt{n}$, the target norm bound on $B^T B$ is $\widetilde{O}(n)$, giving a bound of $\widetilde{O}\left(\frac{1}{n}\right)$ on $\|VRx\|_4^4$.
- If $d \geq \sqrt{n}$, the target norm bound on $B^T B$ is $\widetilde{O}\left(d^2\right)$, giving a bound of $\widetilde{O}\left(\frac{d^2}{n^2}\right)$ on $\|VRx\|_4^4$

Casework



Norm $\tilde{O}(d\sqrt{n})$ if $d \leq \sqrt{n}$, norm $\tilde{O}(d^2)$ if $d \geq \sqrt{n}$

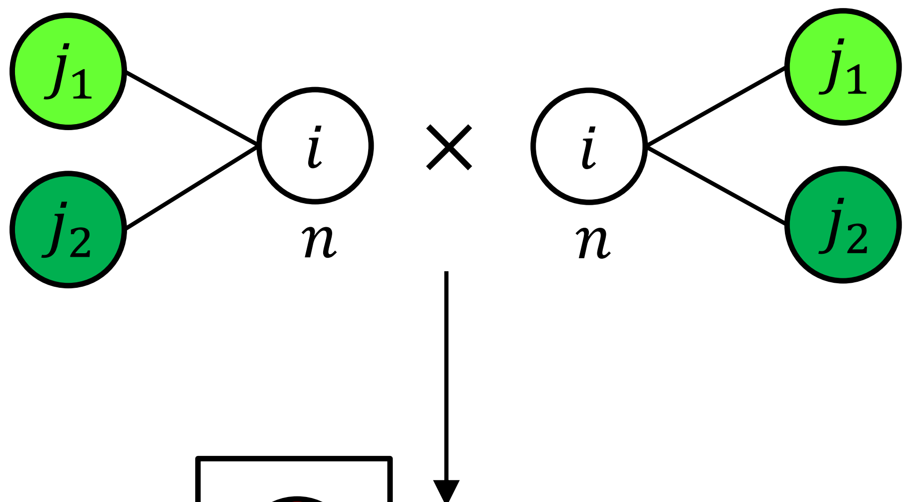
Casework



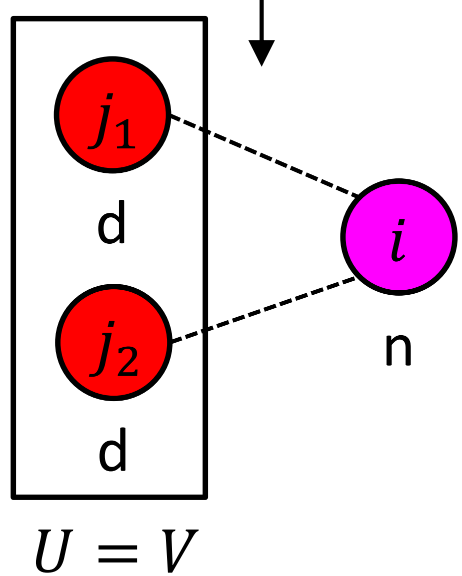
Note: 0 or 2 edges between i and j_1

Norm $\tilde{O}\left(\sqrt{dn}\right)$

Casework

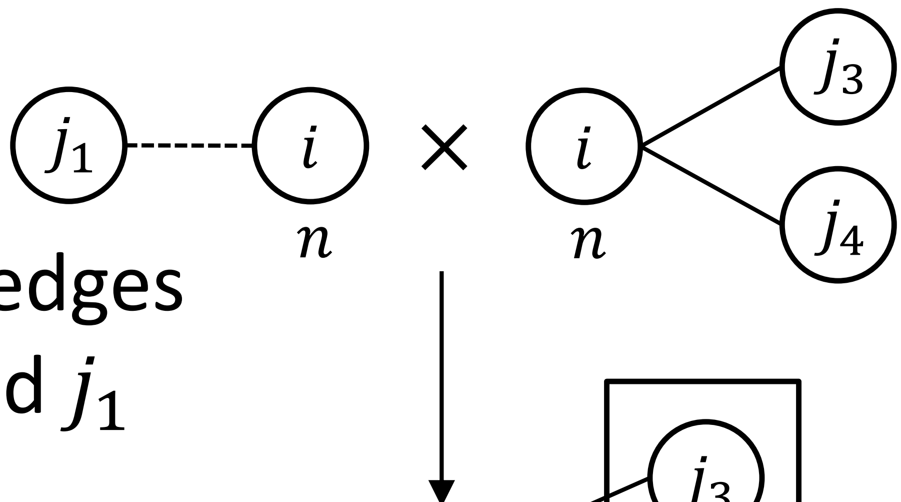


Note: 0 or 2 edges between i and j_1, 0 or 2 edges between i and j_2

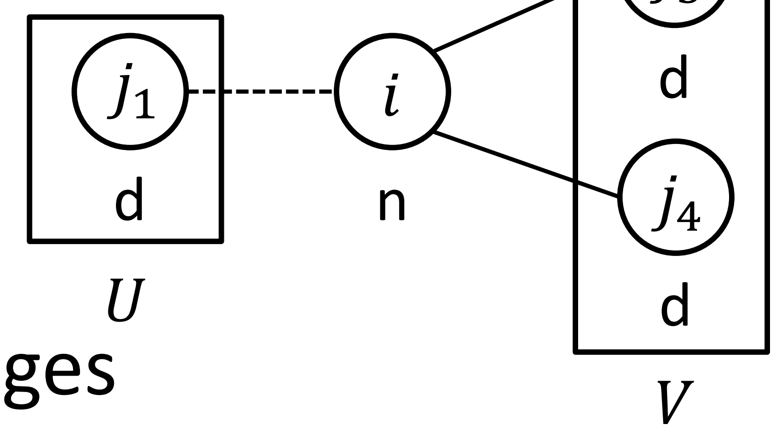


$= nId$ + Norm $\tilde{O}(\sqrt{n})$

Casework



Note: 0 or 2 edges between i and j_1

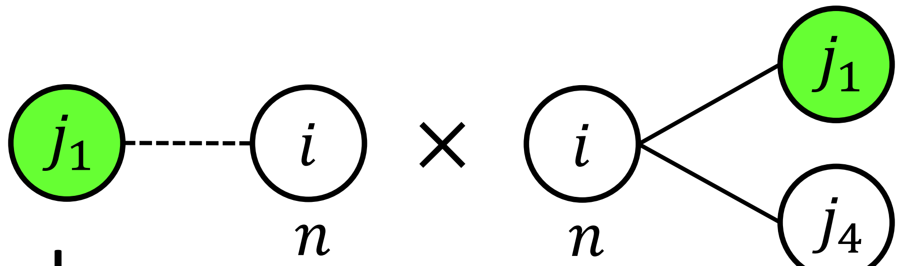


Norm $\tilde{O}\left(\sqrt{nd^3}\right)$

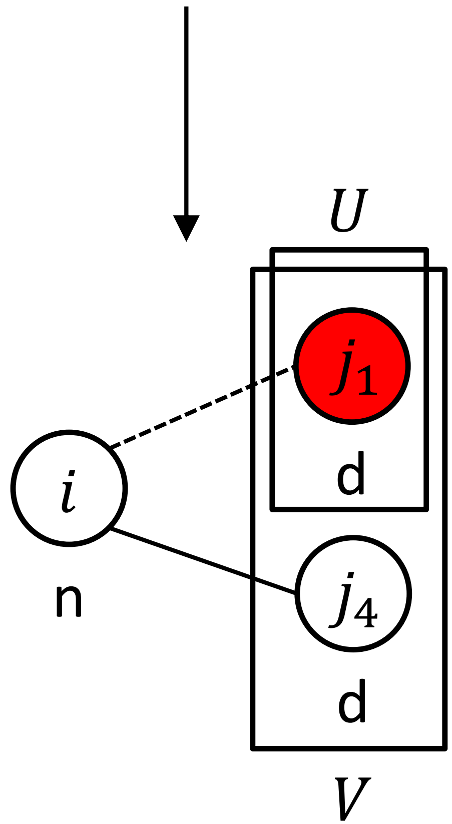
Too large!

Note: 0 or 2 edges between i and j_1

Casework



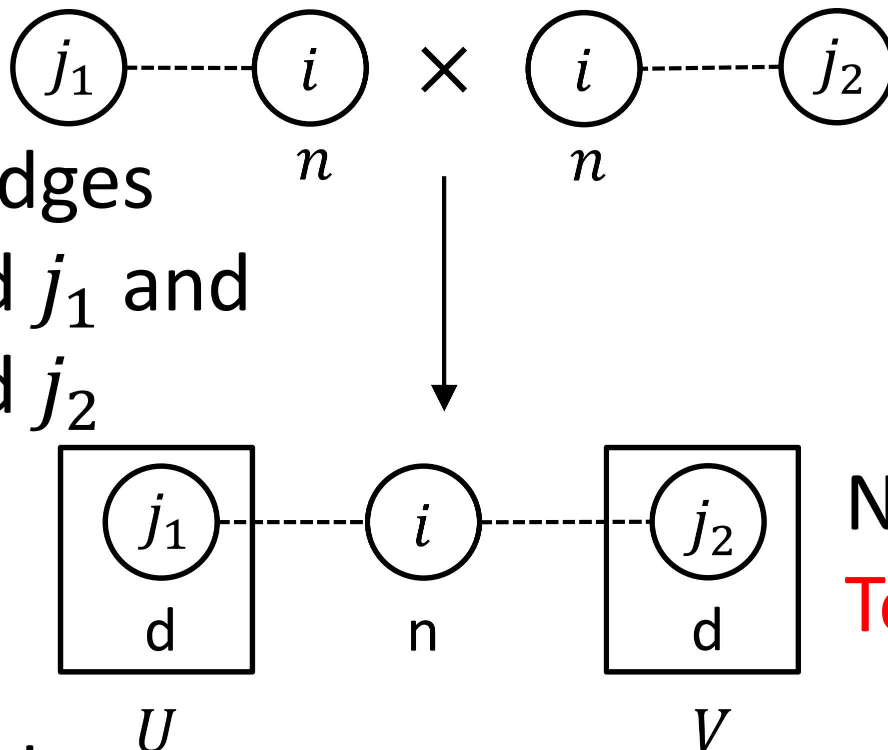
Note: 0 or 2 edges between i and j_1



Note: 1 or 3 edges between i and j_1

Norm $\tilde{O}\left(\sqrt{dn}\right)$

Casework



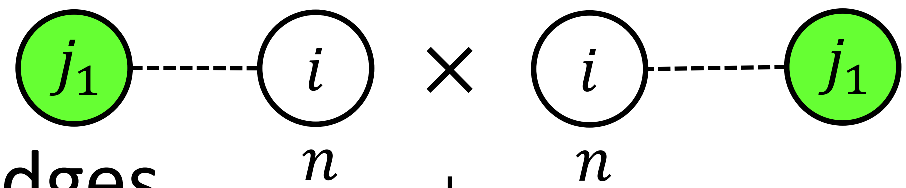
Note: 0 or 2 edges between i and j_1 and between i and j_2

Norm $\tilde{O}(nd)$

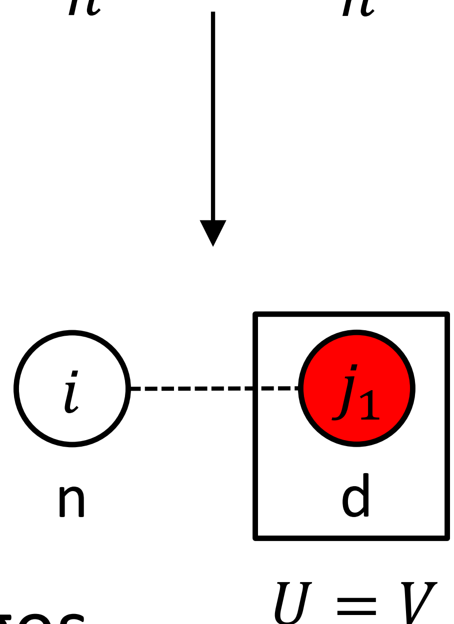
Too large!

Note: 0 or 2 edges between i and j_1 and between i and j_2

Casework



Note: 0 or 2 edges between i and j_1 on both ends



Turns out to be $3Id$ + Norm $\tilde{O}(\sqrt{n})$

Note: 0,2, or 4 edges between i and j_1

Summary

- Most cases have sufficiently small norm.
- Two cases have a norm which is too large, so norm bounds alone are not enough…

Part V: SOS-Symmetry to the Rescue

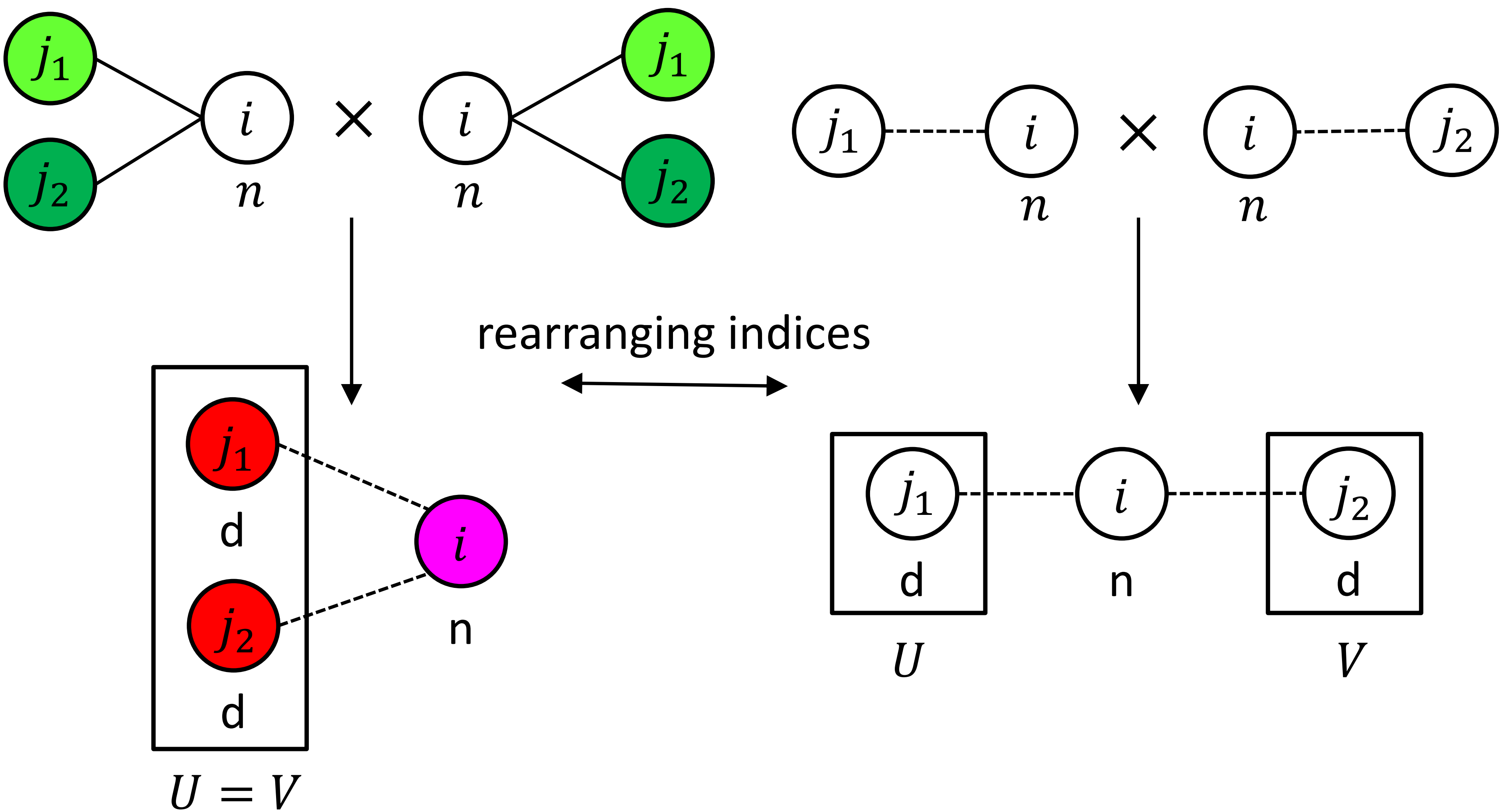
Key Idea: Rearranging Indices

- Instead of looking at $\max_{w:\|w\|=1} w^T B^T B w$, we only need to upper bound

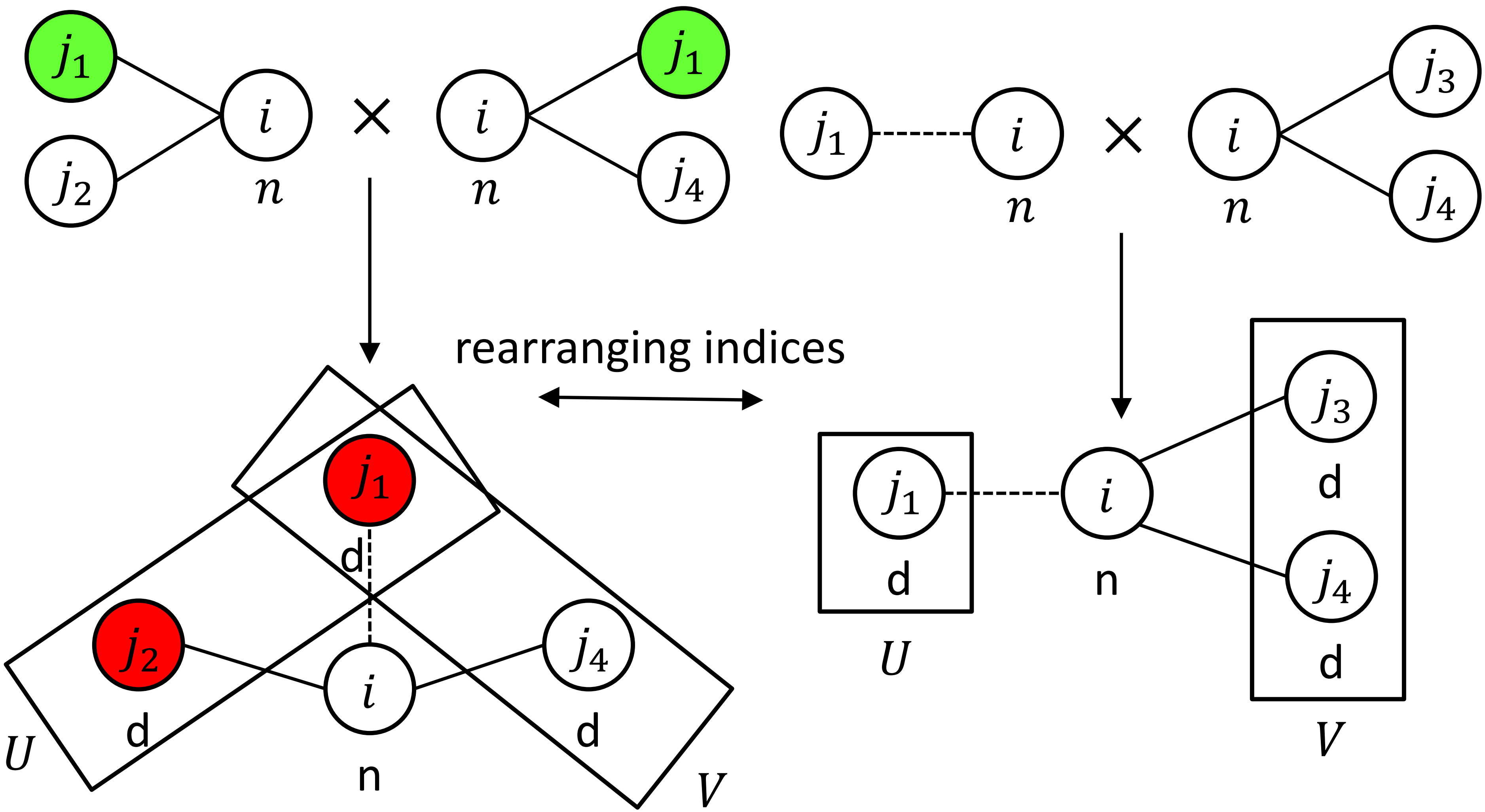
$$\max_{x:\|x\|=1} (x \otimes x)^T B^T B (x \otimes x)$$

- As far as $(x \otimes x)^T B^T B (x \otimes x)$ is concerned, we can rearrange indices in pieces of $B^T B$.

Rearranging Indices Case #1



Rearranging Indices Case #2



Effect of Rearranging Indices

- For the two cases whose norm is too high, their norm can be reduced by rearranging indices.
- This proves the upper bound on

$$\max_{x:\|x\|=1} (x \otimes x)^T B^T B (x \otimes x)$$

Part VI: Observations and Loose Ends

Observations: 4-Norm Analysis

- Note: This 4-norm analysis roughly corresponds to p.33-37 of [BBH+12]
- Remark: When $d \ll \sqrt{n}$, with a slightly more careful analysis we can show that $(x \otimes x)^T B^T B(x \otimes x) = \left(3 \pm o(1)\right)\|x\|_2^4$, matching the results in [BBH+12].

Loose Ends: Arbitrary R

- How can we handle arbitrary R rather than a random orthogonal R (i.e. any span of the vectors)?
- SOS handles it automatically!
- Idea: The SOS-symmetry and $M \succcurlyeq 0$ constraints are invariant under linear transformations of the variables. Thus, having a different R merely applies a linear transformation to the pseudo-expectation values.

Loose Ends: Finding v Exactly

- We have only shown a distinguishing algorithm between the random and planted cases. How can we find the planted sparse vector v exactly?
- Can be done in two steps:
  1. The analysis shows that degree 4 SOS will output a vector v' which is highly correlated with v (because the random part of the subspace has nothing with high 4-norm)
  2. Using v' as a guide, find v. This can be done by minimizing then L^1 norm of a vector v in the subspace subject to $v \cdot v' = 1$, see [BKS14] for details.

Part VII: Open Problems

Open Problems

- What more can we say when $d \gg \sqrt{n}$?
- More specifically, can we find a better algorithm using more than the 4-norm? Is there an SOS lower bound showing that $k = \frac{n^2}{d^2}$ is tight?

References

- [BBH+12] B. Barak, F. G. S. L. Brandão, A. W. Harrow, J. A. Kelner, D. Steurer, and Y. Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. STOC p. 307–326, 2012.
- [BKS14] B. Barak, J. A. Kelner, and D. Steurer. Rounding Sum of Squares Relaxations. STOC 2014.